Data Analysis in R of Blood Pathogens in Alaskan Crows

Abstract

The current COVID-19 pandemic has brought to light the importance of studying infections in animals. While the origins of COVID-19 are still under research and discussion, previous infections in humans originated in animals. Therefore, learning more about infections in various animals and how to identify them becomes a human and global health issue. While there are millions of possibilities on where to research animal infections, there has been research done on crow populations in Alaska observing blood parasites and beak deformities. This project seeks to explore correlations between characteristics of the studied birds and the results of their blood tests. The project will explore geographic significance to see if there are any factors outside of the individual birds that may cause infections. Furthermore, machine learning algorithms will be applied to the datasets to see if any additional patterns or trends can be identified. R is the programming language used through all stages of the data analysis.

Intro/Background of the Problem

Understanding blood pathogens in animals is a key aspect of understanding how virus infections can cross species and possibly into humans. Blood samples, biological data (ex: mass, sex), and observations of beak deformities (AKD) were taken from over 180 crows in Alaska and the blood samples were tested to see if they contain any of the following pathogens: Leucocytozoon parasite infection (LEUC), Haemoproteus parasite infection (HAEM), and Plasmodium parasite infection. The origin of the data collection was initiated by an observation that there was an increase in beak deformities in the Alaskan bird population. Avian Keratin Disorder (AKD) is a beak overgrowth that can cause difficulty preening, feeding, and can lead to high rates of mortality.[1] The study that instigated this data sought to explore if there was a connection between the three infections and AKD.[1]

Data Source

The data was collected from 2007 through 2008 and is administered by the Alaska Science Center. The datasets have been backed-up by the United States Geological Survey. There are two datasets. The first is a catalog of the birds which includes the individual bird's date of record, location, sex, age, classification of beak deformity, tarsus length in millimeters, wing span, and mass. The second dataset includes individual DNA extractions and the results from various tests that were run on these samples.

The data and data dictionary can be accessed at the following link: https://alaska.usgs.gov/products/data.php?dataid=241

Data Preparation

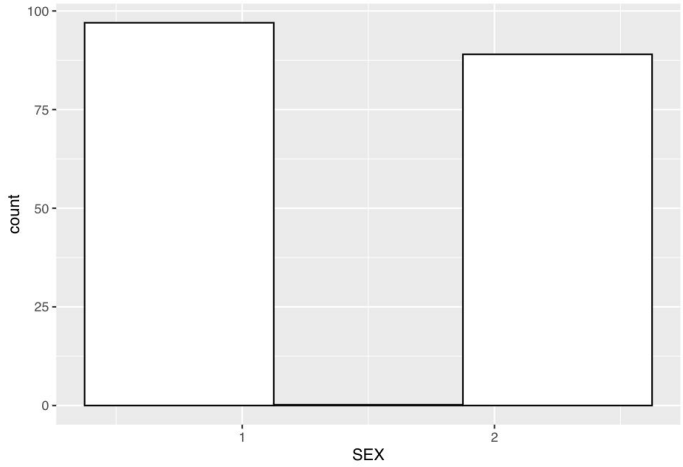
There were three major stages to the data preparation. First, the two datasets needed to be merged based on row number. The second was performing one-hot encoding on the different locations where the samples were collected as this would be better for the machine learning algorithms.

The third stage is how to handle the test results. Each sample was tested twice for each pathogen. For example, the variable LEUC1 shows either a negative or positive result in the first test in a blood sample for Leucocytozoon. The variable LEUC2 shows the same outcome but for a second test on the same sample. As seen in the data exploration, there are some inconsistent results. Overall, the approach for this project will be to use the results from the first test for the analysis of this project.

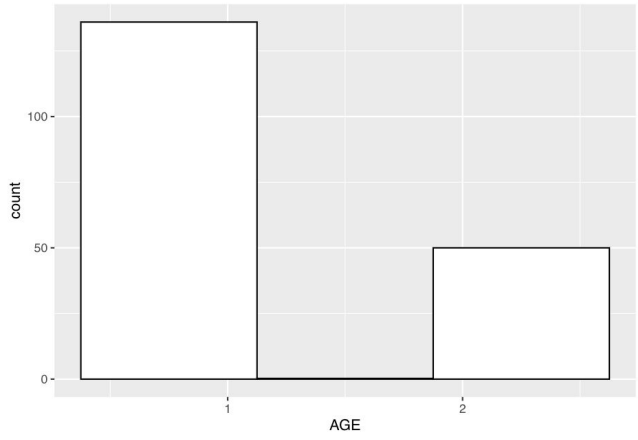
Data Exploration

<u>Variable Distributions</u>

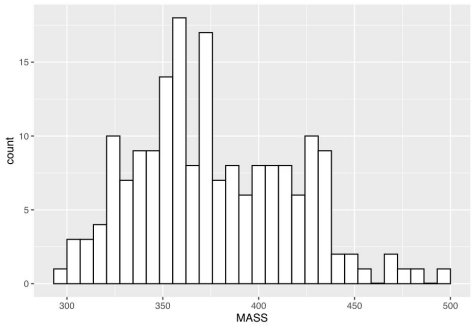
First, it was important to explore the distribution of the variables to see if there were any potential issues to be aware of when analyzing the data. Starting with the biological data, there was a fairly even distribution with sex across samples.

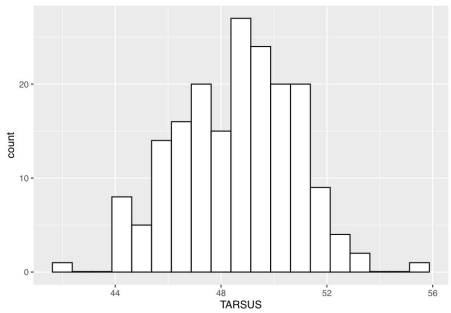


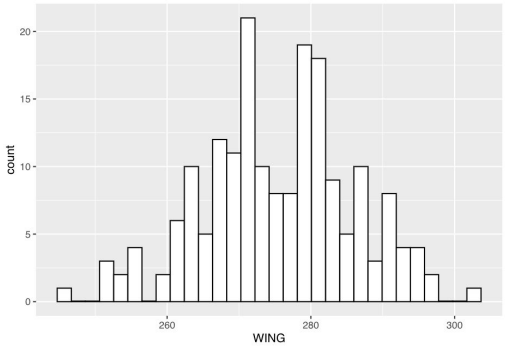
Age is a binary variable of the samples of either one or two years old and most of the samples were deemed to be one year old.



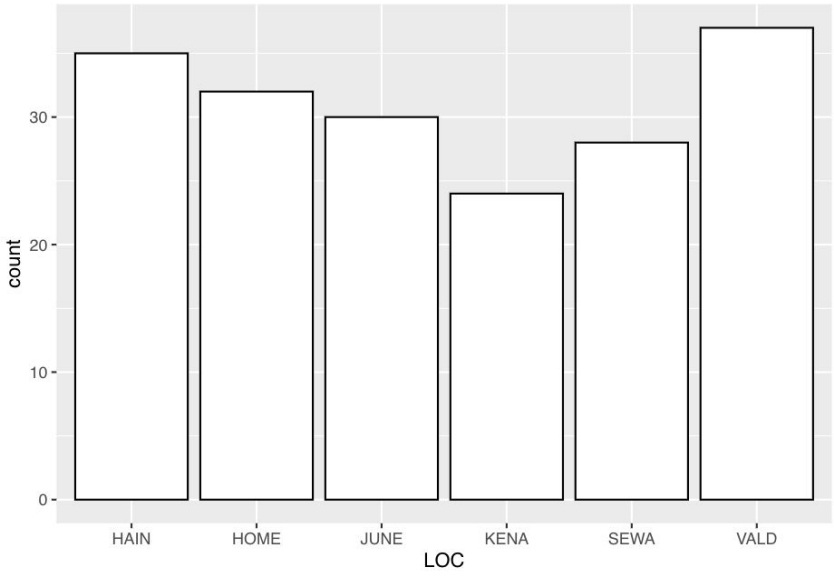
The remaining biological variables are tarsus, wing, and mass. Generally, there is a normal distribution for tarsus and wing but there is a positive skew for the mass data.

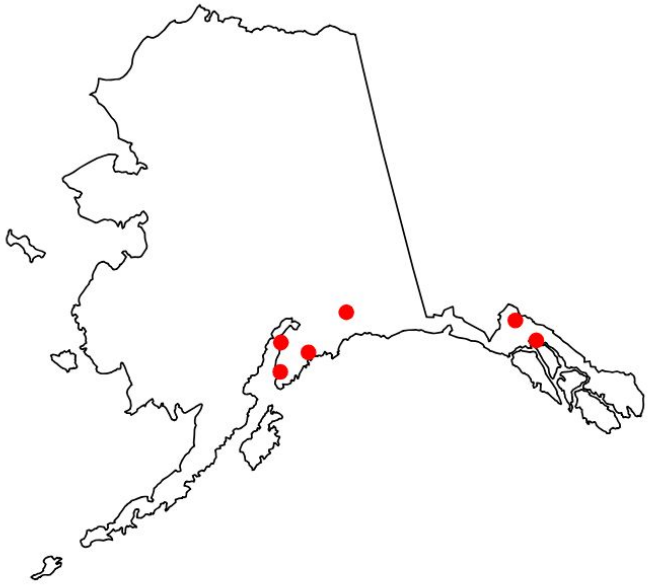




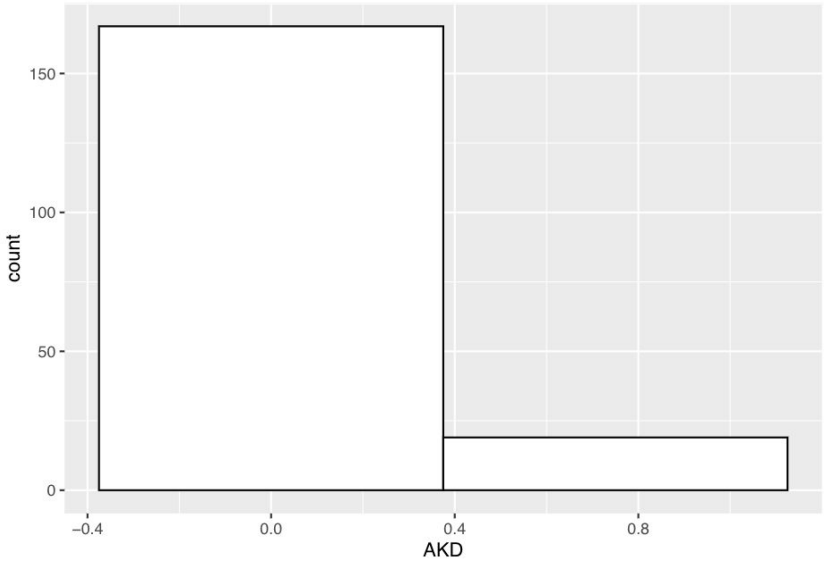


When looking at the locations where the samples were taken, there is a generally even distribution of all six locations (Southcentral: Seward, Kenai, Homer, Valdez. Southeast: Haines, Juneau). However, it may be important to note that Kenai had the smallest amount of samples while Valdez has the largest. The images below show the distribution as well as the geography of the locations that were studied.



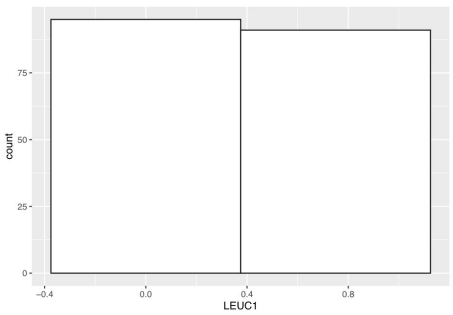


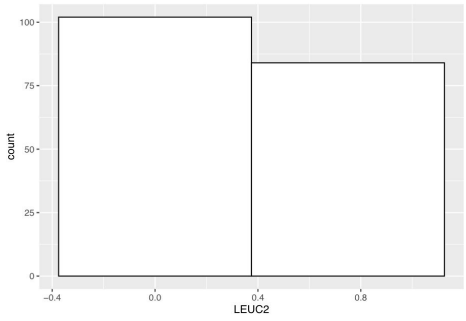
It is also important to note that the majority of the samples collected showed visible signs of AKD.

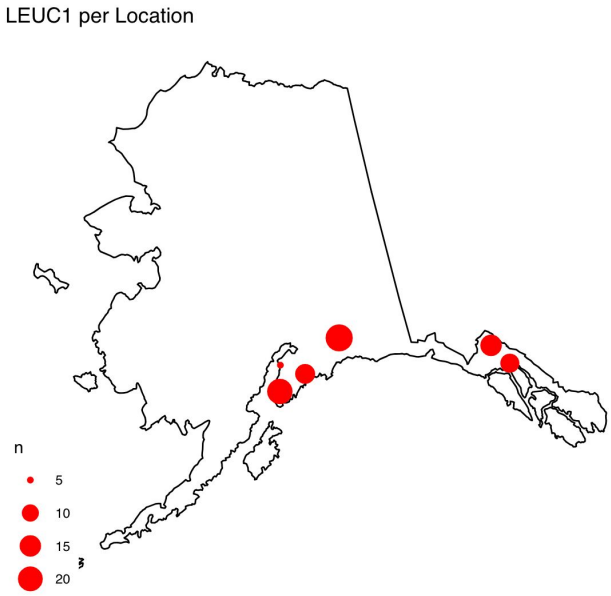


Finally, it was important to see the distribution of the presence of the blood pathogens. For all results, a zero meant negative and a one meant positive. Furthermore, two tests were run on each blood sample.

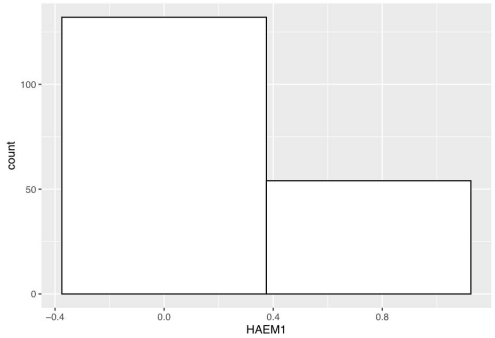
For the LEUC1 and LEUC2 results, the first test showed an approximately even number of positive and negative results while the second test showed more negative results than positive. This does indicate a possible inconsistency in the testnig data (a sample had both a positive and negative result).

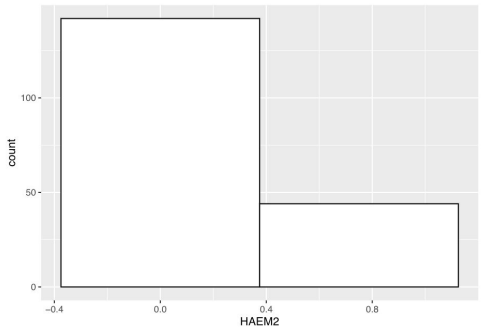


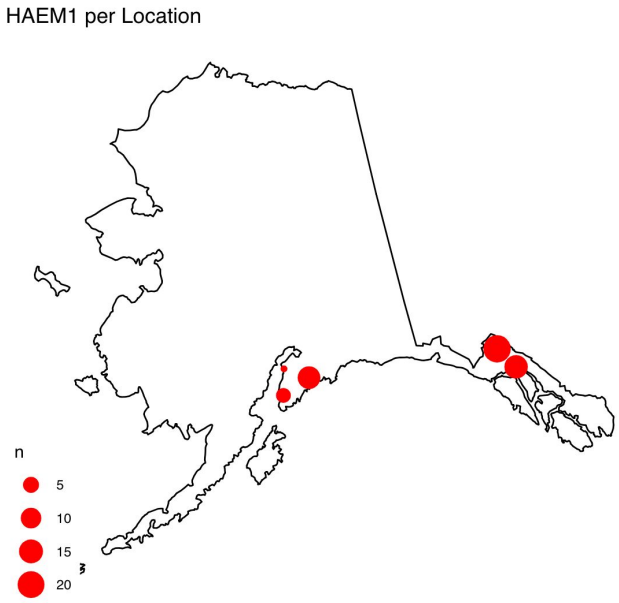




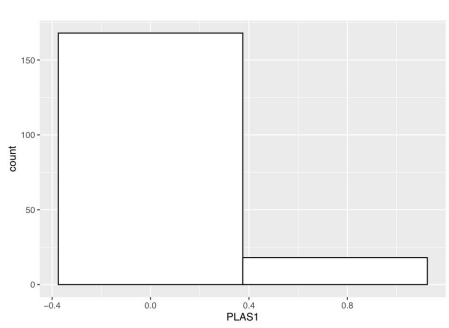
For the HAEM1 and HAEM2 results, there was consistently a higher number of negative responses than positive. However, as with the LEUC1 and LEUC2 results, there appears to be an inconsistency in the testing (a sample had both a positive and negative result).

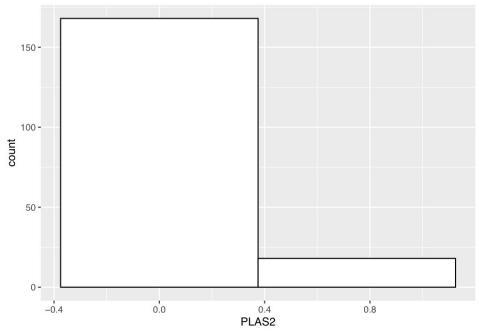


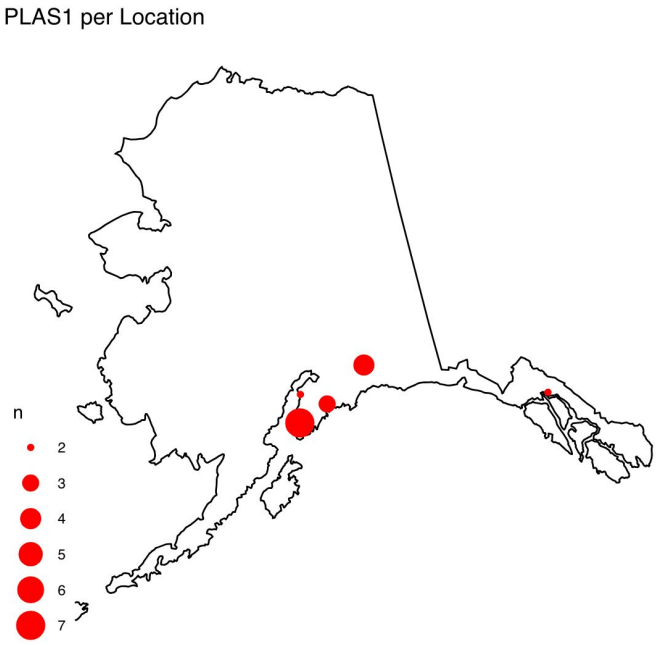




For the PLAS1 and PLAS2 results, there was a very high number of negative results for this test.







Correlations and Covariances

In order to find potential trends and relationships between the variables, the correlation and covariance were calculated with each variable against AKD and the first test results of each blood pathogen. The results showed that there were no significant relationships here.

AKD	Covariance	Correlation Coefficient
Sex	-0.00494	-0.003248
Age	-0.0276	-0.2045
Tarsus	-0.0045	-0.0067
Wing	0.08399	0.0266
LEUC1	-0.0124	-0.0815
HAEM1	-0.003	-0.0202
PLAS1	0.006	0.0697

LEUC1	Covariance	Correlation Coefficient
Sex	-0.0408	-0.1624
Age	-0.0079	-0.0355
Tarsus	0.0926	0.0836
Wing	0.677	0.13001
AKD	-0.0124	-0.0815

HAEM1	Covariance	Correlation Coefficient
Sex	-0.0099	-0.044
Age	0.0405	0.1999
Tarsus	0.1583	0.1574
Wing	-0.2779	-0.05875
AKD	-0.003	-0.0202

PLAS1	Covariance	Correlation Coefficient
Sex	0.0021	0.0141
Age	-0.0153	-0.1164
Tarsus	-0.051	-0.0779
Wing	0.2767	0.0898
AKD	0.006	0.0697

Linear Regression Models

Linear Regression models were created to see if a group of variables possessed any trends or patterns against either AKD or the first test results of each of the blood pathogens. There were no significant relationships found in these models.

Results for Linear Regression on AKD:

```
Call:
lm(formula = AKD ~ LOC + SEX + AGE + PLAS1 + TARSUS + WING +
    MASS + LEUC1 + HAEM1, data = de)

Residuals:
     Min       1Q   Median       3Q      Max
-0.36546 -0.13484 -0.08611  0.02357  0.88740

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.2871459  0.8221499   1.566 0.119316
LOCHOME     -0.0462883  0.0848134  -0.546 0.585946
LOCJUNE      0.0642843  0.0747478   0.860 0.390998
LOCKENA      0.2118082  0.0881992   2.401 0.017414 *
LOCSEWA      0.1922974  0.0766769   2.508 0.013088 *
LOCVALD      0.0643373  0.0807391   0.797 0.426654
SEX         -0.0289801  0.0471944  -0.614 0.540003
AGE         -0.2036639  0.0547149  -3.722 0.000269 ***
PLAS1        0.0704807  0.0773107   0.912 0.363249
TARSUS      -0.0167144  0.0146667  -1.140 0.256059
WING        -0.0018654  0.0034495  -0.541 0.589371


MASS         0.0009558  0.0008629   1.108 0.269584
LEUC1       -0.0203032  0.0470769  -0.431 0.666817
HAEM1        0.0533980  0.0592547   0.901 0.368785
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2941 on 169 degrees of freedom
  (3 observations deleted due to missingness)
Multiple R-squared:  0.1416, Adjusted R-squared:  0.07555
F-statistic: 2.144 on 13 and 169 DF,  p-value: 0.01401
```

Results for Linear Regression on LEUC1:

```
Call:
lm(formula = LEUC1 ~ LOC + SEX + AGE + AKD + TARSUS + WING +
    MASS + HAEM1 + PLAS1, data = de)

Residuals:
     Min       1Q   Median       3Q      Max
-0.89251 -0.42311 -0.07707  0.41033  0.84272

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.099247   1.349700  -0.814   0.4165
LOCHOME      0.308064   0.136590   2.255   0.0254 *
LOCJUNE     -0.030989   0.122313  -0.253   0.8003
LOCKENA     -0.142251   0.146065  -0.974   0.3315
LOCSEWA      0.070881   0.127413   0.556   0.5787
LOCVALD      0.282357   0.130304   2.167   0.0316 *
SEX         -0.135654   0.076450  -1.774   0.0778 .
AGE          0.037018   0.092901   0.398   0.6908
AKD         -0.054148   0.125553  -0.431   0.6668
TARSUS       0.008862   0.024034   0.369   0.7128
WING         0.006698   0.005615   1.193   0.2346
MASS        -0.001738   0.001408  -1.234   0.2187
HAEM1        0.130380   0.096481   1.351   0.1784
PLAS1       -0.054109   0.126497  -0.428   0.6694
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4803 on 169 degrees of freedom
  (3 observations deleted due to missingness)
Multiple R-squared:  0.1477, Adjusted R-squared:  0.08214
F-statistic: 2.253 on 13 and 169 DF,  p-value: 0.009435
```

Results for Linear Regression on HAEM1:

```
Call:
lm(formula = HAEM1 ~ LOC + SEX + AGE + AKD + TARSUS + WING +
    MASS + LEUC1 + PLAS1, data = de)

Residuals:
     Min       1Q   Median       3Q      Max
-0.76967 -0.21705 -0.01256  0.31141  0.92067

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.1754749  1.0723495  -0.164   0.8702
LOCHOME     -0.4012597  0.1055140  -3.803   0.0002 ***
LOCJUNE     -0.1377530  0.0964347  -1.428   0.1550
LOCKENA     -0.5727444  0.1074771  -5.329 3.12e-07 ***
LOCSEWA     -0.1495668  0.1004763  -1.489   0.1385
LOCVALD     -0.6059372  0.0938184  -6.459 1.08e-09 ***
SEX         -0.0120490  0.0611812  -0.197   0.8441
AGE          0.1742929  0.0724775   2.405   0.0173 *
AKD          0.0895593  0.0993822   0.901   0.3688
TARSUS       0.0214752  0.0189956   1.131   0.2599
WING         0.0002679  0.0044711   0.060   0.9523
MASS        -0.0015340  0.0011154  -1.375   0.1708
LEUC1        0.0819928  0.0606744   1.351   0.1784
PLAS1       -0.2020122  0.0991584  -2.037   0.0432 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3809 on 169 degrees of freedom
  (3 observations deleted due to missingness)
Multiple R-squared:  0.3489, Adjusted R-squared:  0.2988
F-statistic: 6.965 on 13 and 169 DF,  p-value: 1.051e-10
```

Results for Linear Regression on PLAS1:

```
Call:
lm(formula = PLAS1 ~ LOC + SEX + AGE + AKD + TARSUS + WING +
    MASS + LEUC1 + HAEM1, data = de)

Residuals:
     Min       1Q   Median       3Q      Max
-0.32626 -0.14261 -0.08166  0.00574  0.90057

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.7122827  0.8200915  -0.869   0.3863
LOCHOME      0.1283414  0.0836752   1.534   0.1269
LOCJUNE      0.0283492  0.0743210   0.381   0.7034
LOCKENA     -0.0376134  0.0889761  -0.423   0.6730
LOCSEWA      0.0898453  0.0772000   1.164   0.2461
LOCVALD      0.0044108  0.0802872   0.055   0.9563
SEX          0.0049820  0.0468934   0.106   0.9155
AGE          0.0099216  0.0564844   0.176   0.8608
AKD          0.0694342  0.0761628   0.912   0.3632
TARSUS      -0.0118939  0.0145846  -0.816   0.4159
WING         0.0067232  0.0033875   1.985   0.0488 *
MASS        -0.0012821  0.0008539  -1.501   0.1351
LEUC1       -0.0199871  0.0467265  -0.428   0.6694
HAEM1       -0.1186574  0.0582434  -2.037   0.0432 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2919 on 169 degrees of freedom
  (3 observations deleted due to missingness)
Multiple R-squared:  0.1128, Adjusted R-squared:  0.0445
F-statistic: 1.652 on 13 and 169 DF,  p-value: 0.07565
```

Machine Learning Models

The target variables being used for the machine learning models are AKD, LEUC1, HAEM1, and PLAS1. The reason for that is the scope of the project focusing on finding what features in a bird can predict these four characteristics. When the target variable is not being predicted, it is included as a feature variable. For example, if the target variable is AKD, then LEUC1, HAEM1, and PLAS1 are included in the list of feature variables. As a reminder, the locations were one-hot encoded for the machine learning models.

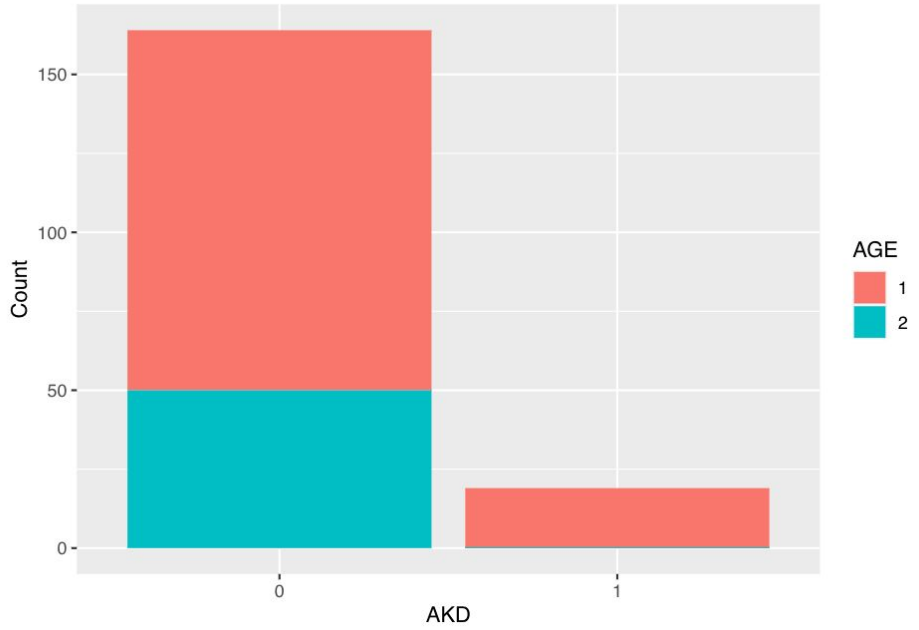
Logistic Regression

The first machine learning model created was a Logistic Regression Model. When predicting AKD, it earned a precision score of 89.78%. The results of the model are below and the most significant revelation was the possibility of a relationship between AKD and Age.

```
##                    2.5 %       97.5 %
## (Intercept) -2634.1900120 2.703235e+03
## SEX            -1.7216634 6.205950e-01
## AGE         -2687.5585861 2.649684e+03
## TARSUS         -0.5595384 1.466764e-01
## WING           -0.1411362 3.639016e-02
## MASS           -0.0105768 4.295700e-02
## LEUC1          -1.4942911 9.954247e-01
## HAEM1          -0.4354375 2.801334e+00
## PLAS1          -1.1808293 2.272975e+00
## LOC.SEWA        0.3159412 5.206111e+00
## LOC.KENA        0.4843588 5.526163e+00
## LOC.VALD       -0.9485564 3.915921e+00
## LOC.HAIN       -3.1540757 2.962598e+00
## LOC.JUNE       -1.3598649 3.762584e+00
## LOC.HOME              NA           NA
```

	FALSE	TRUE
0	121	2
1	12	2

The graph below shows that the only samples that did not show signs of AKD were estimated to be one year old. This reveals a possibility that the signs of AKD do not appear until the bird is at least two years old or that it takes that long for any infection that causes AKD to take effect.



On the LEUC1 variable, the model scored a 69.57% accuracy and for the HAEM1 variable it scored an 85.5% accuracy. The results for these models are below, respectively.

```
##                   2.5 %      97.5 %
## (Intercept) -17.71263917  5.909124006
## SEX          -1.31020091  0.040505299
## AGE          -0.64902989  0.988229952
## TARSUS       -0.16748996  0.251269057
## WING         -0.01898793  0.079701079
## MASS         -0.02030441  0.004506129
## AKD          -1.39307227  0.850198909
## HAEM1        -0.25512635  1.438850338
## PLAS1        -1.41951384  0.862883216
## LOC.SEWA     -2.33247532  0.224501978
## LOC.KENA     -3.49483430 -0.744937835
## LOC.VALD     -1.24105373  0.962057262
## LOC.HAIN     -2.60165698 -0.162969019
## LOC.JUNE     -2.73548874 -0.310925938
## LOC.HOME             NA           NA
```

	FALSE	TRUE
0	49	21
1	21	47

```
##                    2.5 %       97.5 %
## (Intercept) -2.555382e+01 6.897946e+00
## SEX         -9.989220e-01 7.764953e-01
## AGE          2.875900e-01 2.583641e+00
## TARSUS      -1.322820e-01 4.189151e-01
## WING        -5.266542e-02 7.420183e-02
## MASS        -2.597683e-02 5.088832e-03
## LEUC1       -2.017605e-01 1.558561e+00
## AKD         -8.411244e-01 2.107398e+00
## PLAS1       -4.370732e+03 4.334582e+03
## LOC.SEWA    -5.049756e-01 2.711085e+00
## LOC.KENA    -4.645821e+00 2.663481e-01
## LOC.VALD    -3.216350e+03 3.179455e+03
## LOC.HAIN     3.676849e-01 3.202834e+00
## LOC.JUNE    -1.759188e-01 2.749479e+00
## LOC.HOME              NA           NA
```

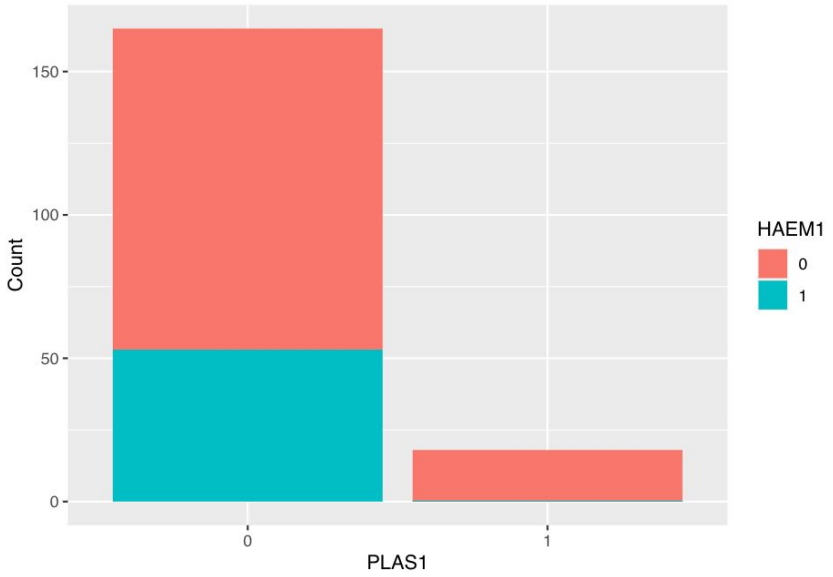
	FALSE	TRUE
0	90	8
1	12	28

The Logistic Regression Model on PLAS1 scored a 92% accuracy and highlighted a potential relationship between PLAS1 and HAEM1 as well as a potential relationship between PLAS1 and the region of Haines.

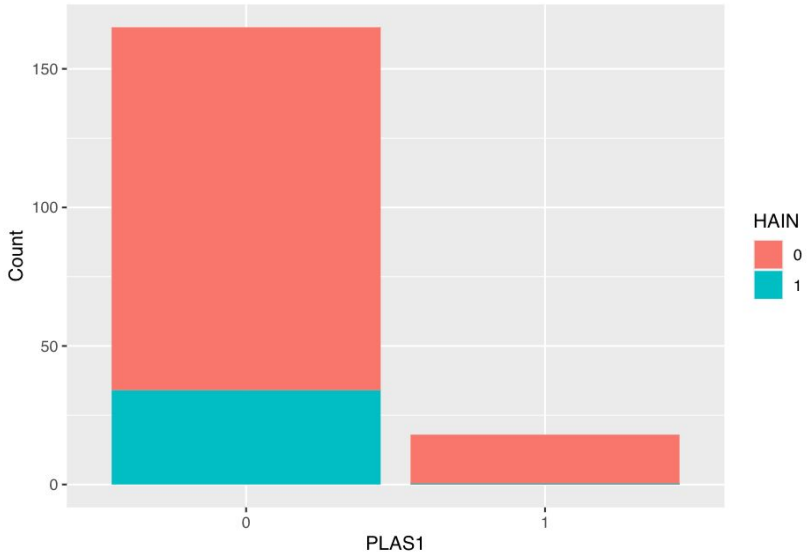
```
##                    2.5 %       97.5 %
## (Intercept) -4.024072e+01  3.899895e+00
## SEX         -1.182815e+00  1.132212e+00
## AGE         -1.785850e+00  1.738433e+00
## TARSUS      -5.059899e-01  2.209062e-01
## WING         2.125323e-02  2.226448e-01
## MASS        -5.219369e-02 -1.072603e-03
## LEUC1       -1.421506e+00  9.861846e-01
## HAEM1       -4.066102e+03  4.028546e+03
## AKD         -7.635654e-01  2.683554e+00
## LOC.SEWA    -1.478637e+00  2.524080e+00
## LOC.KENA    -3.550206e+00  5.054281e-01
## LOC.VALD    -2.356527e+00  5.392387e-01
## LOC.HAIN    -4.900745e+03  4.865228e+03
## LOC.JUNE    -2.369083e+00  1.309431e+00
## LOC.HOME              NA           NA
```

	FALSE	TRUE
0	123	1
1	10	4

The below graph shows that for the significant majority of positive results for HAEM1, the sample had a negative result of PLAS1.



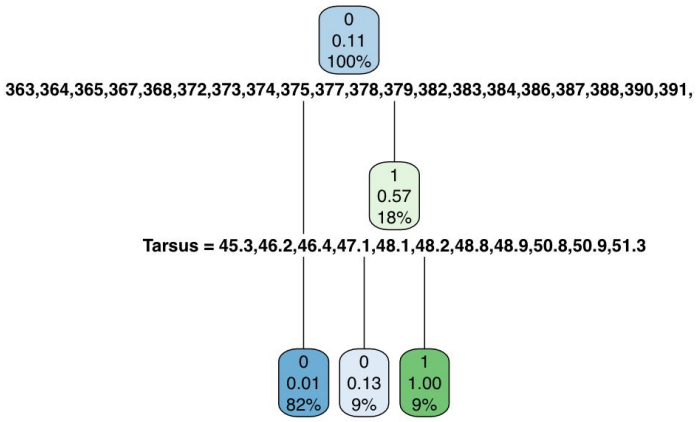
The below graph shows that none of the positive results for PLAS1 came from the region of Haines. However, it is important to recall that most of the samples collected across all regions returned a negative result for PLAS1.



Decision Tree

The second machine learning model used on the data was a decision tree model. There are two main reasons for this. The first reason is that all target variables are binary which is suited for the decision tree. The second reason is due to the lack of results found when exploring covariance, correlation, and linear regression. A decision tree can display what feature variables can help make decisions in getting to the target variable which can help with further understanding of the relationships between the variables.

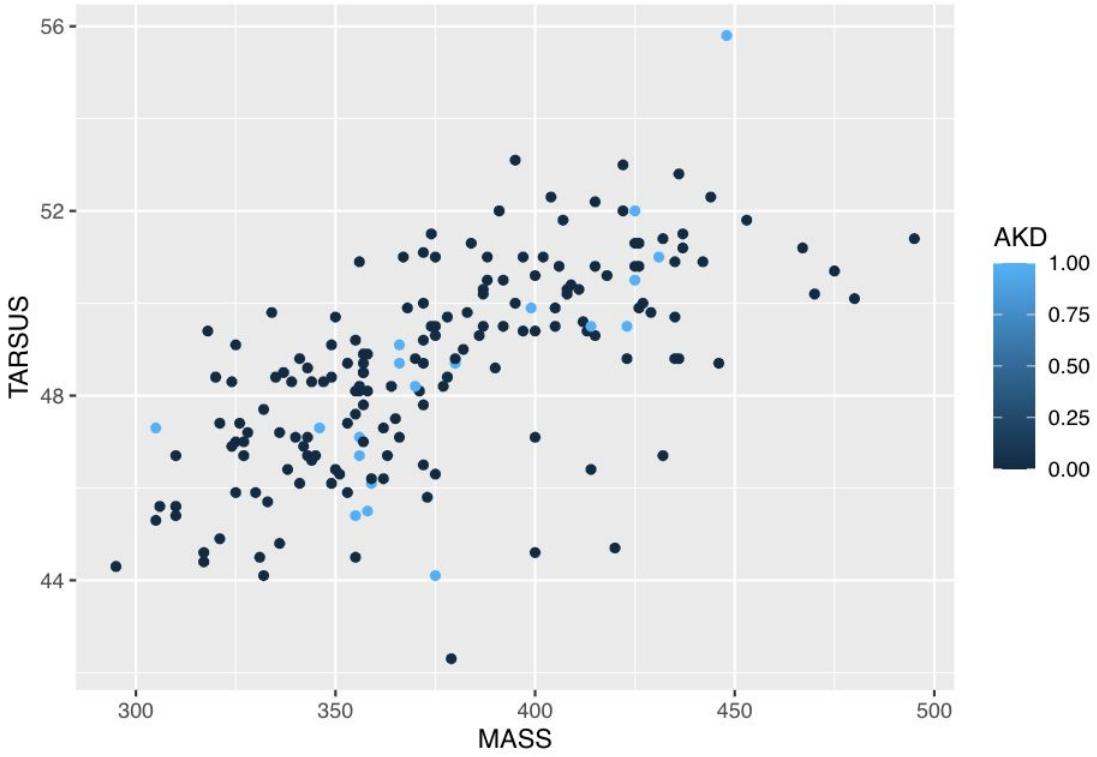
The decision tree predicting AKD was able to get a 94.7% accuracy using the Mass and Tarsus variables.




```
p <- predict(dtm, data_test, type="class")
confMat <- table(data_test$AKD,p)
accuracy <- sum(diag(confMat))/sum(confMat)
return (accuracy*100)
```

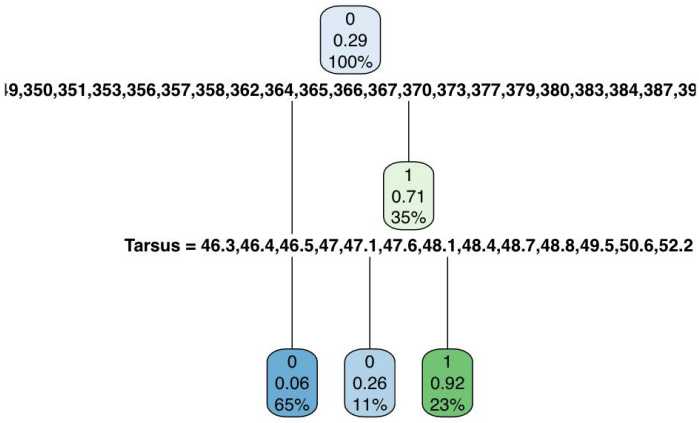
```
## [1] 94.73684
```

For further exploration, the below scatterplot shows the Tarsus and Mass variables plotted against each other with the color being decided by the presence or absence of AKD. Blue (1) means the sample was negative for AKD and Black (0) means the sample was positive for AKD. A significant majority of the samples were positive for AKD. There is no visible pattern as to what the relationship between the three variables may be but the Decision Tree model was able to use these data to make a prediction with high accuracy.



The decision tree models for the blood pathogens were not successful and the scores ranged from 47.37% to 73.68%. Many variables were used but the model was unable to find a path to a high precision score. The plotted Decision Trees are included below.

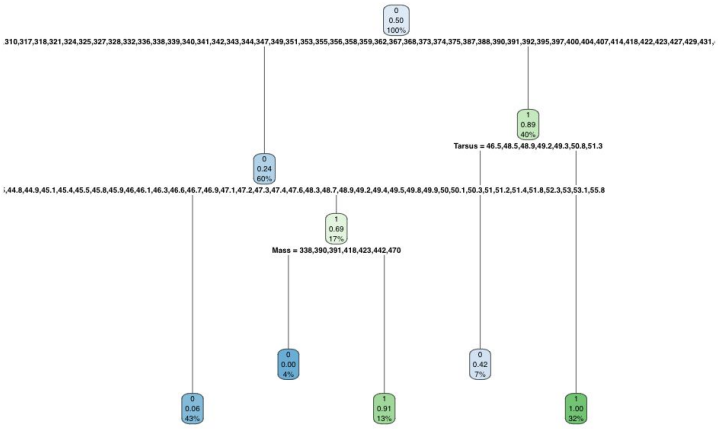
Decision Tree for HAEM1:




```
p <- predict(dtm, data_test, type="class")
confMat <- table(data_test$HAEM1,p)
accuracy <- sum(diag(confMat))/sum(confMat)
return (accuracy*100)
```

```
## [1] 63.15789
```

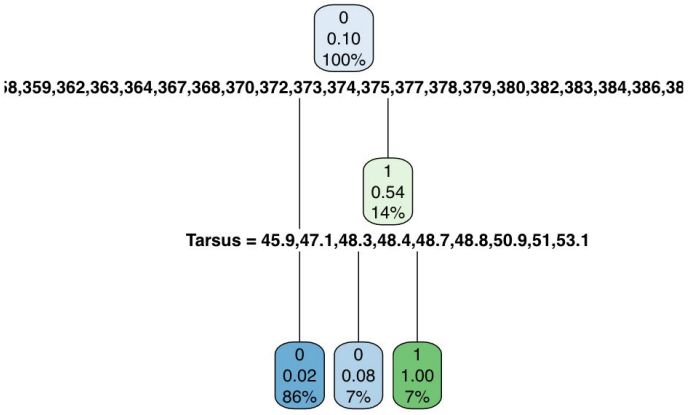
Decision Tree for LEUC1:




```
p <- predict(dtm, data_test, type="class")
confMat <- table(data_test$LEUC1,p)
accuracy <- sum(diag(confMat))/sum(confMat)
return (accuracy*100)
```

```
## [1] 47.36842
```

Decision Tree for PLAS1:




```
p <- predict(dtm, data_test, type="class")
confMat <- table(data_test$PLAS1,p)
accuracy <- sum(diag(confMat))/sum(confMat)
return (accuracy*100)
```

```
## [1] 73.68421
```

Naive-Bayes

The third machine learning algorithm used was Naive Bayes. When using AKD as the target variable, the Naive Bayes model was able to get 87.5%. However, since the majority of samples were positive for AKD, when run on the test sample, the Naive Bayes model predicted that all samples were positive for AKD and the 87.5% happened to be the ratio of the test sample that was positive for AKD. Overall, this model did not analyze the data to find the few samples that were not positive for AKD.

For the blood pathogens, the Naive Bayes model scored 50% predicting LEUC1, 75% predicting HAEM1, and 78.125% predicting PLAS1.

The results for all machine learning algorithms and the target variables are included in the table below.

	Target Variable			
Algorithm	**AKD**	**LEUC1**	**HAEM1**	**PLAS1**
Logistic Regression	89.78%	69.57%	85.50%	92%
Decision Tree	94.70%	47.37%	63.16%	73.68%
Naive Bayes	87.50%	50%	75%	78.13%

Results and Conclusion

While some extrapolations from relationships between the variables were able to be identified, there were no relationships or trends that were significant enough to make predictions on answering such questions about outbreaks in certain locations, biological details that could predict the presence of pathogens, or the presence of pathogens predicting the presence or details of other variables. However, the Logistic Regression machine learning model showed potential for a model to be created to make predictions on certain features. The Logistic Regression model scored high accuracy for AKD, HAEM1, and PLAS1. The Decision Tree held the most promise for predicting the presence of AKD and highlighted a possible relationship between mass and tarsus and AKD while the Logistic Regression model signalled an importance of age on the sample. The Naive Bayes model was only able to provide a significant accuracy for predicting AKD but it failed to see the nuances in the few samples that were negative for AKD.

References

[1]Van Hemert, C., Meixell, B.W., Smith, M.M., and Handel, C.M. (2019 Jun 10). "Prevalence and diversity of avian blood parasites in a resident northern passerine." Retrieved from https://parasitesandvectors.biomedcentral.com/articles/10.1186/s13071-019-3545-1

While the data I am exploring is from 2007 and 2008, this article provides a recent exploration of blood parasites in northern birds. This will be able to provide further domain context into how to understand the data I am using. The research is focused on beak deformities in Northern birds and the impact of climate.

[2]Oakgrove, K., Harrigan, R., Loiseau, C., Guers, S., Seppi, B., & Sehgal, R. (2014). Distribution, diversity and drivers of blood-borne parasite co-infections in Alaskan bird populations. International Journal for Parasitology. 44. 10.1016/j.ijpara.2014.04.011.

A study of environmental determinants of the prevalence, diversity, and co-infections of four avian blood-borne parasites in Alaskan bird populations.

[3]Fecchio, A., Bell, J., Bosholn, M., Vaughan, J., Tkach, V.,l Lutz, H., Cueto, V., Gorosito, C., González-Acuña, D., Stromlund, C., Kvasager, D., Comiche, K., Kirchgatter, K., & Pinho, J., & Berv, J., Anciaes, M., Fontana, C., Zyskowski, K., & Sampaio, S., & Clark, N. (2020). An inverse latitudinal gradient in infection probability and phylogenetic diversity for Leucocytozoon blood parasites in New World birds. Journal of Animal Ecology. 89. 423–435. 10.1111/1365-2656.13117.

Research into how geographical variation and environmental conditions that promote parasite transmission. The project explored sixty-nine bird communities and how characteristics of the hosts and their ecology impacted the growth of an avian blood parasite.

[4]Fecchio, A., & Chagas, C., & Bell, J., & Kirchgatter, K. (2020). Evolutionary ecology, taxonomy, and systematics of avian malaria and related parasites. Acta Tropica. 204. 105364. 10.1016/j.actatropica.2020.105364.

Research on the most common parasites infecting birds, how they have impacted human diseases such as malaria, and the impact on spread from birds' migratory patterns.

[5]Cuevas, E., Vianna, J., Botero-Delgadillo, E., Doussang, D., González-Acuña, D., Barroso, O., Rozzi, R., Vásquez, R., & Quirici, V. (2019). Latitudinal gradients of haemosporidian parasites: Prevalence, diversity and drivers of infection in the Thorn-tailed Rayadito (Aphrastura spinicauda). International Journal for Parasitology: Parasites and Wildlife. 11. 10.1016/j.ijppaw.2019.11.002.

Research exploring how latitudinal gradients impact the infection and spread of haemosporidian parasites in birds.

[6](2015 Feb 10). "Transmission of Avian Influenza A Viruses Between Animals and People." Retrieved from https://www.cdc.gov/flu/avianflu/virus-transmission.htm

Information from the CDC on how the Avian Flu spread from birds to humans.

[7]Branswell, H. (2019 Feb 13). "What happened to bird flu? How a major threat to human health faded from view." Retrieved from https://www.statnews.com/2019/02/13/bird-flu-mutations-outlook/

A retrospective on the bird flu and how it spread amongst birds and a look at what could happen next if certain factors happen.

[8]Handell, C., Van Hemert, C. R., Pajot, L.M., and Richardson, R.M. "Physical Description of Beak Deformities." Retrieved from https://www.usgs.gov/centers/asc/science/physical-description-beak-deformities?qt-science_center_objects=0#qt-science_center_objects

An overview provided by the Alaska Science Center on the various beak deformities that are found in Alaskan birds and what they may indicate about that bird's health.

[9]"Birds and their droppings can carry over 60 diseases." Retrieved on April 12th, 2020 from https://www.medicalnewstoday.com/releases/61646#1

An article looking at the various diseases that spread from birds to humans and how they are transmitted.

[10](2018 Oct 20). “How can birds make you sick?” Retrieved from https://www.rentokil.ie/blog/how-can-birds-make-you-sick/

An exploration from the Rentokil Pest Control Ireland organization on the range of diseases in birds and how they can spread to humans.